

RAW SEQUENCE LISTING

The Biotechnology Systems Branch of the Scientific and Technical Information Center (STIC) no errors detected.

Application Serial Number: 10/563,572
Source: TPWP
Date Processed by STIC: F19-06

ENTERED



IFWP

RAW SEQUENCE LISTING

DATE: 01/19/2006

PATENT APPLICATION: US/10/563,572

TIME: 13:29:52

Input Set : A:\GFAT-anglais.txt

Output Set: N:\CRF4\01192006\J563572.raw

4 <110> APPLICANT: CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE

6 <120> TITLE OF INVENTION: GLUTAMINE:FRUCTOSE-6-PHOSPHATE AMIDOTRANSFERASE (GFAT)

COMPRISING

7 AN INTERNAL PURIFICATION TAG, AND ITS USE FOR THE SCREENING
 8 OF COMPOUNDS

10 <130> FILE REFERENCE: WOB 03 BP CNR GFAT

C--> 12 <140> CURRENT APPLICATION NUMBER: US/10/563,572

C--> 12 <141> CURRENT FILING DATE: 2006-01-06

12 <160> NUMBER OF SEQ ID NOS: 19

14 <170> SOFTWARE: PatentIn version 3.1

16 <210> SEQ ID NO: 1

17 <211> LENGTH: 2046

18 <212> TYPE: DNA

19 <213> ORGANISM: Homo sapiens

21 <220> FEATURE:

22 <221> NAME/KEY: CDS

23 <222> LOCATION: (1)..(2046)

24 <223> OTHER INFORMATION:

26 <220> FEATURE:

27 <221> NAME/KEY: misc_feature

28 <222> LOCATION: (170)..(170)

29 <223> OTHER INFORMATION: t ou c

W--> 32 <400> 1

33 atg tgt ggt ata ttt gct tac tta aac tac cat gtt cct cga acg aga 48

34 Met Cys Gly Ile Phe Ala Tyr Leu Asn Tyr His Val Pro Arg Thr Arg

35 1 5 10 15

37 cga gaa atc ctg gag acc cta atc aaa ggc ctt cag aga ctg gag tac 96

38 Arg Glu Ile Leu Glu Thr Leu Ile Lys Gly Leu Gln Arg Leu Glu Tyr

39 20 25 30

41 aga gga tat gat tct gct ggt gtg gga ttt gat gga ggc aat gat aaa 144

42 Arg Gly Tyr Asp Ser Ala Gly Val Gly Phe Asp Gly Gly Asn Asp Lys

43 35 40 45

W--> 45 gat tgg gaa gcc aat gcc tgc aaa anc cag ctt att aag aag aaa gga 192

W--> 46 Asp Trp Glu Ala Asn Ala Cys Lys Xaa Gln Leu Ile Lys Lys Lys Gly

47 50 55 60

49 aaa gtt aag gca ctg gat gaa gaa gtt cac aag caa caa gat atg gat 240

50 Lys Val Lys Ala Leu Asp Glu Glu Val His Lys Gln Gln Asp Met Asp

51 65 70 75 80

53 ttg gat ata gaa ttt gat gta cac ctt gga ata gct cat acc cgt tgg 288

54 Leu Asp Ile Glu Phe Asp Val His Leu Gly Ile Ala His Thr Arg Trp

55 85 90 95

57 gca aca cat gga gaa ccc agt cct gtc aat agc cac ccc cag cgc tct 336

58 Ala Thr His Gly Glu Pro Ser Pro Val Asn Ser His Pro Gln Arg Ser

59 100 105 110

RAW SEQUENCE LISTING

DATE: 01/19/2006

PATENT APPLICATION: US/10/563,572

TIME: 13:29:52

Input Set : A:\GFAT-anglais.txt

Output Set: N:\CRF4\01192006\J563572.raw

63	gat	aaa	aat	aat	gaa	ttt	atc	gtt	att	cac	aat	gga	atc	atc	acc	aac	384
64	Asp	Lys	Asn	Asn	Glu	Phe	Ile	Val	Ile	His	Asn	Gly	Ile	Ile	Thr	Asn	
65			115					120					125				
67	tac	aaa	gac	ttg	aaa	aag	ttt	ttg	gaa	agc	aaa	ggc	tat	gac	ttc	gaa	432
68	Tyr	Lys	Asp	Leu	Lys	Lys	Phe	Leu	Glu	Ser	Lys	Gly	Tyr	Asp	Phe	Glu	
69			130				135					140					
71	tct	gaa	aca	gac	aca	gag	aca	att	gcc	aag	ctc	gtt	aag	tat	atg	tat	480
72	Ser	Glu	Thr	Asp	Thr	Glu	Thr	Ile	Ala	Lys	Leu	Val	Lys	Tyr	Met	Tyr	
73	145						150					155				160	
75	gac	aat	cgg	gaa	agt	caa	gat	acc	agc	ttt	act	acc	ttg	gtg	gag	aga	528
76	Asp	Asn	Arg	Glu	Ser	Gln	Asp	Thr	Ser	Phe	Thr	Thr	Leu	Val	Glu	Arg	
77						165				170					175		
79	gtt	atc	caa	caa	ttg	gaa	ggt	gct	ttt	gca	ctt	gtg	ttt	aaa	agt	gtt	576
80	Val	Ile	Gln	Gln	Leu	Glu	Gly	Ala	Phe	Ala	Leu	Val	Phe	Lys	Ser	Val	
81				180						185				190			
83	cat	ttt	ccc	ggg	caa	gca	gtt	ggc	aca	agg	cga	ggt	agc	cct	ctg	ttg	624
84	His	Phe	Pro	Gly	Gln	Ala	Val	Gly	Thr	Arg	Arg	Gly	Ser	Pro	Leu	Leu	
85			195					200					205				
87	att	ggt	gta	cgg	agt	gaa	cat	aaa	ctt	tct	act	gat	cac	att	cct	ata	672
88	Ile	Gly	Val	Arg	Ser	Glu	His	Lys	Leu	Ser	Thr	Asp	His	Ile	Pro	Ile	
89		210					215					220					
91	ctc	tac	aga	aca	ggc	aaa	gac	aag	aaa	gga	agc	tgc	aat	ctc	tct	cgt	720
92	Leu	Tyr	Arg	Thr	Gly	Lys	Asp	Lys	Lys	Gly	Ser	Cys	Asn	Leu	Ser	Arg	
93	225					230					235				240		
95	gtg	gac	agc	aca	acc	tgc	ctt	ttc	ccg	gtg	gaa	gaa	aaa	gca	gtg	gag	768
96	Val	Asp	Ser	Thr	Thr	Cys	Leu	Phe	Pro	Val	Glu	Glu	Lys	Ala	Val	Glu	
97				245						250				255			
99	tat	tac	ttt	gct	tct	gat	gca	agt	gct	gtc	ata	gaa	cac	acc	aat	cgc	816
100	Tyr	Tyr	Phe	Ala	Ser	Asp	Ala	Ser	Ala	Val	Ile	Glu	His	Thr	Asn	Arg	
101			260							265				270			
103	gtc	atc	ttt	ctg	gaa	gat	gat	gat	gtt	gca	gca	gta	gtg	gat	gga	cgt	864
104	Val	Ile	Phe	Leu	Glu	Asp	Asp	Asp	Val	Ala	Ala	Val	Val	Asp	Gly	Arg	
105			275					280					285				
107	ctt	tct	atc	cat	cga	att	aaa	cga	act	gca	gga	gat	cac	ccc	gga	cga	912
108	Leu	Ser	Ile	His	Arg	Ile	Lys	Arg	Thr	Ala	Gly	Asp	His	Pro	Gly	Arg	
109		290					295					300					
111	gct	gtg	caa	aca	ctc	cag	atg	gaa	ctc	cag	cag	atc	atg	aag	ggc	aac	960
112	Ala	Val	Gln	Thr	Leu	Gln	Met	Glu	Leu	Gln	Gln	Ile	Met	Lys	Gly	Asn	
113	305					310					315				320		
115	ttc	agt	tca	ttt	atg	cag	aag	gaa	ata	ttt	gag	cag	cca	gag	tct	gtc	1008
116	Phe	Ser	Ser	Phe	Met	Gln	Lys	Glu	Ile	Phe	Glu	Gln	Pro	Glu	Ser	Val	
117				325						330				335			
119	gtg	aac	aca	atg	aga	gga	aga	gtc	aac	ttt	gat	gac	tat	act	gtg	aat	1056
120	Val	Asn	Thr	Met	Arg	Gly	Arg	Val	Asn	Phe	Asp	Asp	Tyr	Thr	Val	Asn	
121			340							345				350			
125	ttg	ggt	ggt	ttg	aag	gat	cac	ata	aag	gag	atc	cag	aga	tgc	cgg	cgt	1104
126	Leu	Gly	Gly	Leu	Lys	Asp	His	Ile	Lys	Glu	Ile	Gln	Arg	Cys	Arg	Arg	
127			355					360					365				
129	ttg	att	ctt	att	gct	tgt	gga	aca	agt	tac	cat	gct	ggt	gta	gca	aca	1152

RAW SEQUENCE LISTING

DATE: 01/19/2006

PATENT APPLICATION: US/10/563,572

TIME: 13:29:52

Input Set : A:\GFAT-anglais.txt

Output Set: N:\CRF4\01192006\J563572.raw

```

130 Leu Ile Leu Ile Ala Cys Gly Thr Ser Tyr His Ala Gly Val Ala Thr
131      370                      375                      380
133 cgt caa gtt ctt gag gag ctg act gag ttg cct gtg atg gtg gaa cta      1200
134 Arg Gln Val Leu Glu Glu Leu Thr Glu Leu Pro Val Met Val Glu Leu
135 385                      390                      395                      400
137 gca agt gac ttc ctg gac aga aac aca cca gtc ttt cga gat gat gtt      1248
138 Ala Ser Asp Phe Leu Asp Arg Asn Thr Pro Val Phe Arg Asp Asp Val
139                      405                      410                      415
141 tgc ttt ttc ctt agt caa tca ggt gag aca gca gat act ttg atg ggt      1296
142 Cys Phe Phe Leu Ser Gln Ser Gly Glu Thr Ala Asp Thr Leu Met Gly
143                      420                      425                      430
145 ctt cgt tac tgt aag gag aga gga gct tta act gtg ggg atc aca aac      1344
146 Leu Arg Tyr Cys Lys Glu Arg Gly Ala Leu Thr Val Gly Ile Thr Asn
147                      435                      440                      445
149 aca gtt ggc agt tcc ata tca cgg gag aca gat tgt gga gtt cat att      1392
150 Thr Val Gly Ser Ser Ile Ser Arg Glu Thr Asp Cys Gly Val His Ile
151                      450                      455                      460
153 aat gct ggt cct gag att ggt gtg gcc agt aca aag gct tat acc agc      1440
154 Asn Ala Gly Pro Glu Ile Gly Val Ala Ser Thr Lys Ala Tyr Thr Ser
155 465                      470                      475                      480
157 cag ttt gta tcc ctt gtg atg ttt gcc ctt atg atg tgt gat gat cgg      1488
158 Gln Phe Val Ser Leu Val Met Phe Ala Leu Met Met Cys Asp Asp Arg
159                      485                      490                      495
161 atc tcc atg caa gaa aga cgc aaa gag atc atg ctt gga ttg aaa cgg      1536
162 Ile Ser Met Gln Glu Arg Arg Lys Glu Ile Met Leu Gly Leu Lys Arg
163                      500                      505                      510
165 ctg cct gat ttg att aag gaa gta ctg agc atg gat gac gaa att cag      1584
166 Leu Pro Asp Leu Ile Lys Glu Val Leu Ser Met Asp Asp Glu Ile Gln
167                      515                      520                      525
169 aaa cta gca aca gaa ctt tat cat cag aag tca gtt ctg ata atg gga      1632
170 Lys Leu Ala Thr Glu Leu Tyr His Gln Lys Ser Val Leu Ile Met Gly
171                      530                      535                      540
173 cga ggc tat cat tat gct act tgt ctt gaa ggg gca ctg aaa atc aaa      1680
174 Arg Gly Tyr His Tyr Ala Thr Cys Leu Glu Gly Ala Leu Lys Ile Lys
175 545                      550                      555                      560
177 gaa att act tat atg cac tct gaa ggc atc ctt gct ggt gaa ttg aaa      1728
178 Glu Ile Thr Tyr Met His Ser Glu Gly Ile Leu Ala Gly Glu Leu Lys
179                      565                      570                      575
181 cat ggc cct ctg gct ttg gtg gat aaa ttg atg cct gtg atc atg atc      1776
182 His Gly Pro Leu Ala Leu Val Asp Lys Leu Met Pro Val Ile Met Ile
183                      580                      585                      590
187 atc atg aga gat cac act tat gcc aag tgt cag aat gct ctt cag caa      1824
188 Ile Met Arg Asp His Thr Tyr Ala Lys Cys Gln Asn Ala Leu Gln Gln
189                      595                      600                      605
191 gtg gtt gct cgg cag ggg cgg cct gtg gta att tgt gat aag gag gat      1872
192 Val Val Ala Arg Gln Gly Arg Pro Val Val Ile Cys Asp Lys Glu Asp
193                      610                      615                      620
195 act gag acc att aag aac aca aaa aga acg atc aag gtg ccc cac tca      1920
196 Thr Glu Thr Ile Lys Asn Thr Lys Arg Thr Ile Lys Val Pro His Ser

```

RAW SEQUENCE LISTING

DATE: 01/19/2006

PATENT APPLICATION: US/10/563,572

TIME: 13:29:52

Input Set : A:\GFAT-anglais.txt

Output Set: N:\CRF4\01192006\J563572.raw

```

197 625          630          635          640
199 gtg gac tgc ttg cag ggc att ctc agc gtg atc cct tta cag ttg ctg      1968
200 Val Asp Cys Leu Gln Gly Ile Leu Ser Val Ile Pro Leu Gln Leu Leu
201          645          650          655
203 gct ttc cac ctt gct gtg ctg aga ggc tat gat gtt gat ttc cca cgg      2016
204 Ala Phe His Leu Ala Val Leu Arg Gly Tyr Asp Val Asp Phe Pro Arg
205          660          665          670
207 aat ctt gcc aaa tct gtg act gta gag tga      2046
208 Asn Leu Ala Lys Ser Val Thr Val Glu
209          675          680
212 <210> SEQ ID NO: 2
213 <211> LENGTH: 681
214 <212> TYPE: PRT
215 <213> ORGANISM: Homo sapiens
217 <220> FEATURE:
218 <221> NAME/KEY: misc_feature
219 <222> LOCATION: (57)..(57)
220 <223> OTHER INFORMATION: 'Xaa' in position 57 represents Thr or Ile.
222 <400> SEQUENCE: 2
223 Met Cys Gly Ile Phe Ala Tyr Leu Asn Tyr His Val Pro Arg Thr Arg
224 1          5          10          15
226 Arg Glu Ile Leu Glu Thr Leu Ile Lys Gly Leu Gln Arg Leu Glu Tyr
227          20          25          30
229 Arg Gly Tyr Asp Ser Ala Gly Val Gly Phe Asp Gly Gly Asn Asp Lys
230          35          40          45
W--> 232 Asp Trp Glu Ala Asn Ala Cys Lys Xaa Gln Leu Ile Lys Lys Lys Gly
233          50          55          60
235 Lys Val Lys Ala Leu Asp Glu Glu Val His Lys Gln Gln Asp Met Asp
236 65          70          75          80
238 Leu Asp Ile Glu Phe Asp Val His Leu Gly Ile Ala His Thr Arg Trp
239          85          90          95
241 Ala Thr His Gly Glu Pro Ser Pro Val Asn Ser His Pro Gln Arg Ser
242          100          105          110
244 Asp Lys Asn Asn Glu Phe Ile Val Ile His Asn Gly Ile Ile Thr Asn
245          115          120          125
248 Tyr Lys Asp Leu Lys Lys Phe Leu Glu Ser Lys Gly Tyr Asp Phe Glu
249          130          135          140
251 Ser Glu Thr Asp Thr Glu Thr Ile Ala Lys Leu Val Lys Tyr Met Tyr
252 145          150          155          160
254 Asp Asn Arg Glu Ser Gln Asp Thr Ser Phe Thr Thr Leu Val Glu Arg
255          165          170          175
257 Val Ile Gln Gln Leu Glu Gly Ala Phe Ala Leu Val Phe Lys Ser Val
258          180          185          190
260 His Phe Pro Gly Gln Ala Val Gly Thr Arg Arg Gly Ser Pro Leu Leu
261          195          200          205
263 Ile Gly Val Arg Ser Glu His Lys Leu Ser Thr Asp His Ile Pro Ile
264          210          215          220
266 Leu Tyr Arg Thr Gly Lys Asp Lys Lys Gly Ser Cys Asn Leu Ser Arg
267 225          230          235          240

```

RAW SEQUENCE LISTING

DATE: 01/19/2006

PATENT APPLICATION: US/10/563,572

TIME: 13:29:52

Input Set : A:\GFAT-anglais.txt

Output Set: N:\CRF4\01192006\J563572.raw

```

269 Val Asp Ser Thr Thr Cys Leu Phe Pro Val Glu Glu Lys Ala Val Glu
270                               245                               250                               255
272 Tyr Tyr Phe Ala Ser Asp Ala Ser Ala Val Ile Glu His Thr Asn Arg
273                               260                               265                               270
275 Val Ile Phe Leu Glu Asp Asp Asp Val Ala Ala Val Val Asp Gly Arg
276                               275                               280                               285
278 Leu Ser Ile His Arg Ile Lys Arg Thr Ala Gly Asp His Pro Gly Arg
279                               290                               295                               300
281 Ala Val Gln Thr Leu Gln Met Glu Leu Gln Gln Ile Met Lys Gly Asn
282 305                               310                               315                               320
284 Phe Ser Ser Phe Met Gln Lys Glu Ile Phe Glu Gln Pro Glu Ser Val
285                               325                               330                               335
287 Val Asn Thr Met Arg Gly Arg Val Asn Phe Asp Asp Tyr Thr Val Asn
288                               340                               345                               350
290 Leu Gly Gly Leu Lys Asp His Ile Lys Glu Ile Gln Arg Cys Arg Arg
291                               355                               360                               365
293 Leu Ile Leu Ile Ala Cys Gly Thr Ser Tyr His Ala Gly Val Ala Thr
294                               370                               375                               380
296 Arg Gln Val Leu Glu Glu Leu Thr Glu Leu Pro Val Met Val Glu Leu
297 385                               390                               395                               400
299 Ala Ser Asp Phe Leu Asp Arg Asn Thr Pro Val Phe Arg Asp Asp Val
300                               405                               410                               415
302 Cys Phe Phe Leu Ser Gln Ser Gly Glu Thr Ala Asp Thr Leu Met Gly
303                               420                               425                               430
305 Leu Arg Tyr Cys Lys Glu Arg Gly Ala Leu Thr Val Gly Ile Thr Asn
306                               435                               440                               445
309 Thr Val Gly Ser Ser Ile Ser Arg Glu Thr Asp Cys Gly Val His Ile
310                               450                               455                               460
312 Asn Ala Gly Pro Glu Ile Gly Val Ala Ser Thr Lys Ala Tyr Thr Ser
313 465                               470                               475                               480
315 Gln Phe Val Ser Leu Val Met Phe Ala Leu Met Met Cys Asp Asp Arg
316                               485                               490                               495
318 Ile Ser Met Gln Glu Arg Arg Lys Glu Ile Met Leu Gly Leu Lys Arg
319                               500                               505                               510
321 Leu Pro Asp Leu Ile Lys Glu Val Leu Ser Met Asp Asp Glu Ile Gln
322                               515                               520                               525
324 Lys Leu Ala Thr Glu Leu Tyr His Gln Lys Ser Val Leu Ile Met Gly
325                               530                               535                               540
327 Arg Gly Tyr His Tyr Ala Thr Cys Leu Glu Gly Ala Leu Lys Ile Lys
328 545                               550                               555                               560
330 Glu Ile Thr Tyr Met His Ser Glu Gly Ile Leu Ala Gly Glu Leu Lys
331                               565                               570                               575
333 His Gly Pro Leu Ala Leu Val Asp Lys Leu Met Pro Val Ile Met Ile
334                               580                               585                               590
336 Ile Met Arg Asp His Thr Tyr Ala Lys Cys Gln Asn Ala Leu Gln Gln
337                               595                               600                               605
339 Val Val Ala Arg Gln Gly Arg Pro Val Val Ile Cys Asp Lys Glu Asp
340                               610                               615                               620
342 Thr Glu Thr Ile Lys Asn Thr Lys Arg Thr Ile Lys Val Pro His Ser

```

RAW SEQUENCE LISTING ERROR SUMMARY
PATENT APPLICATION: US/10/563,572

DATE: 01/19/2006
TIME: 13:29:53

Input Set : A:\GFAT-anglais.txt
Output Set: N:\CRF4\01192006\J563572.raw

Please Note:

Use of n and/or Xaa have been detected in the Sequence Listing. Please review the Sequence Listing to ensure that a corresponding explanation is presented in the <220> to <223> fields of each sequence which presents at least one n or Xaa.

Seq#:1; N Pos. 170 ✓
Seq#:1; Xaa Pos. 57 ✓
Seq#:2; Xaa Pos. 57 ✓
Seq#:5; N Pos. 170
Seq#:5; Xaa Pos. 57
Seq#:6; Xaa Pos. 57
Seq#:7; N Pos. 170
Seq#:7; Xaa Pos. 57
Seq#:8; Xaa Pos. 57
Seq#:11; N Pos. 170
Seq#:11; Xaa Pos. 57
Seq#:12; Xaa Pos. 57

VERIFICATION SUMMARY

DATE: 01/19/2006

PATENT APPLICATION: US/10/563,572

TIME: 13:29:53

Input Set : A:\GFAT-anglais.txt

Output Set: N:\CRF4\01192006\J563572.raw

L:12 M:270 C: Current Application Number differs, Replaced Current Application No
L:12 M:271 C: Current Filing Date differs, Replaced Current Filing Date
L:32 M:258 W: Mandatory Feature missing, <223> Blank for SEQ#:1,Line#:24
L:45 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:1 after pos.:144
L:46 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:1 after pos.:192
L:232 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:2 after pos.:48
L:365 M:258 W: Mandatory Feature missing, <223> Blank for SEQ#:3,Line#:363
L:699 M:258 W: Mandatory Feature missing, <223> Blank for SEQ#:5,Line#:692
L:712 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:5 after pos.:144
L:713 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:5 after pos.:192
L:903 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:6 after pos.:48
L:1050 M:258 W: Mandatory Feature missing, <223> Blank for SEQ#:7,Line#:1040
L:1063 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:7 after pos.:144
L:1064 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:7 after pos.:192
L:1249 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:8 after pos.:48
L:1386 M:258 W: Mandatory Feature missing, <223> Blank for SEQ#:9,Line#:1384
L:1723 M:258 W: Mandatory Feature missing, <223> Blank for SEQ#:11,Line#:1713
L:1736 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:11 after pos.:144
L:1737 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:11 after pos.:192
L:1929 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:12 after pos.:48